Bayesian Inference

Shivaram Kalyanakrishnan

April 3, 2019

Abstract

In this note, we describe the principle behind Bayesian inference and illustrate it with two examples: (1) modeling belief about the bias of a coin that is sequentially tossed, and (2) modeling belief about a changing world that is observed regularly. As an application of Bayesian reasoning, we present Thompson Sampling, which is an optimal sampling algorithm for stochastic multi-armed bandits. We provide a brief description of the algorithm.

1 Bayesian Inference

Bayesian inference or Bayesian reasoning is a rigorous framework to model uncertainty. The Bayesian thinker begins with a *prior* belief about the true world being any one from a possible set of worlds. As evidence (produced by the real, unknown world) becomes available, the Bayesian thinker applies his/her knowledge of the process by which evidence is produced by each possible world—and accordingly refines his/her set of beliefs.

As a concrete example, suppose we have a coin whose bias is unknown to us. It is up to us to have some prior belief about this unknown bias. For example, Janaki might be unwilling to afford any bias a higher chance of being true than any other bias; in other words, Janaki's belief is distributed uniformly over $[0, 1]$, which is the set of all biases. Altamash, on the other hand, might have a more worldly perspective, which leads him to believe that our particular coin must "pretty much" be like other coins encountered in life. One way he may encode his belief is to place it uniformly in, say, $[0.48, 0.52]$.

In general, prior beliefs can be *arbitrary*. However, as we might expect, the closer one's prior belief is to reality, the more accurate one's reasoning based on that belief. In practice, one often uses domain knowledge to encode prior beliefs. As a case in point, consider on-line ads, which can be viewed as coins. When shown to a user (tossed), an ad either gets clicked (shows up heads) or does not get clicked (shows up tails). It is rare for "click-through rates" to exceed 5%. Hence, Mrinalini, who places all of her belief in $[0, 0.05]$, will likely take better decisions about a typical on-line ad than Janaki, who has distributed her belief uniformly in $[0, 1]$. In turn, Janaki will do a lot better than Guru, who restricts his belief to the interval $[0.99, 0.995]$.

Bayesian inference does not stop with having a useful prior belief. At the core of Bayesian reasoning is the operation of *updating* one's belief based on evidence. Regardless of what their initial beliefs are, Mrinalini, Janaki, Altamash, and Guru should presumably have relatively similar beliefs (if they are rational, that is!) after they have seen a thousand independent displays of the ad in question. After all, if the ad had a click-through rate of p, it precisely means that the probability that the ad generates a click is p. If, in a thousand displays, twelve yielded clicks and the remaining tails, shouldn't Guru's belief change significantly from Uniform([0.99, 0.995])? How must Guru update his beliefs based on evidence? By applying Bayes' rule, of course!

2 Belief Distribution

The Belief distribution (or simply *Belief*) is taken to be a distribution over $World$, the random variable denoting possible worlds, given the available evidence. Before any evidence is available (at time 0), we take

$$Belief_0 = P(World),$$

which is nothing but our prior belief. $Belief_0$ will be a discrete probability distribution if the set of possible worlds is discrete. Otherwise, as in our motivating example, wherein each world is a coin bias, $Belief_0$ can be represented using a probability density function (such as Uniform([0, 1])). For every $t \geq 1$, we assume that a new piece of evidence E_t becomes available. Having observed events $E_1, E_2, \ldots, E_t$, we have

$$Belief_t = P(World|E_1, E_2, \ldots, E_t).$$

We assume that evidence is generated sequentially through a process that is determined by the world: that is, pieces of evidence are conditionally independent given the world. This assumption is consistent with the Bayes Net shown in Figure 1.

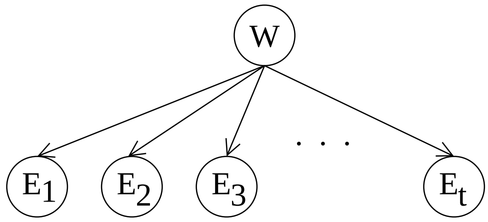

Figure 1: Bayes Net showing each unit of evidence E_t being generated independently by the world W.

The cornerstone of Bayesian inference is the fact that $Belief_t$ is a "sufficient statistic" of the prior and of all the evidence seen thus far—in the sense that $Belief_{t+1}$ can be obtained solely from $Belief_t$ and E_{t+1}. To see how, consider applying Bayes' rule, and arguing that given a world, evidence events are conditionally independent. We have:

$$\begin{aligned} P(W|E_1, E_2, \ldots, E_{t+1}) &= \frac{P(E_1, E_2, \ldots, E_{t+1}|W)P(W)}{P(E_1, E_2, \ldots, E_{t+1})} \\ &= \frac{P(E_1, E_2, \ldots, E_t|W)P(E_{t+1}|W)P(W)}{P(E_1, E_2, \ldots, E_{t+1})} \\ &= \frac{P(E_1, E_2, \ldots, E_t, W)P(E_{t+1}|W)}{P(E_1, E_2, \ldots, E_{t+1})} \\ &= \frac{P(W|E_1, E_2, \ldots, E_t)P(E_1, E_2, \ldots, E_t)P(E_{t+1}|W)}{P(E_1, E_2, \ldots, E_{t+1})} \\ &\propto P(W|E_1, E_2, \ldots, E_t)P(E_{t+1}|W). \end{aligned}$$

In other words, we have

$$Belief_{t+1}(W) \propto Belief_t(W) \cdot P(E_{t+1}|W).$$

Informally, this expression is often described as "posterior is proportional to the product of the prior and the likelihood". The belief distribution is usually represented in some parametric form: whenever evidence becomes available, the parameters of the distribution are updated. We consider a concrete example for illustration.

3 Belief Distribution over Coin Biases

We step into Janaki's shoes. Recall that her prior belief over the set of coins is Uniform([0, 1]). Let $x \in [0, 1]$ be an arbitrary bias; we have (Janaki's)

$$Belief_0(x) = 1 \cdot dx,$$

where she represents belief as a probability density function over the interval $[0, 1]$ (we multiply by the differential dx to ensure that the belief is actually a probability). Now assume that the coin gets tossed, and it shows up heads. We get

$$Belief_1(x) \propto Belief_0(x) P(Head|x) = 1 \cdot dx \cdot x,$$

which yields

$$Belief_1(x) = \frac{x dx}{\int_{y=0}^{1} y dy} = 2x \cdot dx.$$

Suppose the second toss results in a tail. Janaki's belief would now be updated to

$$Belief_2(x) \propto Belief_1(x) P(Tail|x) = 2x \cdot dx \cdot (1 - x),$$

which yields

$$Belief_2(x) = \frac{2x(1 - x) dx}{\int_{y=0}^{1} 2y(1 - y) dy} = 6x(1 - x) \cdot dx.$$

Observe that in keeping with intuition, $Belief_1$ afforded higher probability to biases exceeding 0.5, since the first toss resulted in a head. However, after seeing a head and a tail, the belief ($Belief_2$) is symmetric about 0.5. Would this also be the case had we started with Mrinalini's prior belief instead of Janaki's?

Interestingly, in the case of coin biases (Bernoulli variables whose draws constitute the evidence), beliefs that start as uniform priors, and which are updated according to Bayes' rule, exactly correspond to Beta distributions[1] with parameters α and β, wherein α is the number of observed heads plus one, and β is the number of observed tails plus one. As a sanity check, we must expect $Beta(1, 1)$ to be the uniform distribution, which it is! Another implication is that our posterior belief after seeing h heads and t tails will be the same regardless of the actual sequence in which the heads and tails were generated. This property also seems reasonable: after all, we know that for every given bias, sequences that contain the same numbers of heads and tails have the same probability of being generated.

It is convenient in the coin-bias world that if the prior belief is a Beta distribution, then so is the posterior belief. Hence, all that needs to be done to keep an up-to-date belief distribution is to keep count of the number of heads and tails! However, it is not generally true that the prior and the posterior distribution will be from the same parametric family (that is, they need not be *conjugate*).

4 Case Study: Thompson Sampling

What is the point of keeping a belief distribution over the set of possible worlds, and refining it based on evidence?! An agent that has to *act* in an unknown world can potentially use its belief distribution to good effect. Indeed there is a very relevant learning problem to solve with coins, the domain we have used as a running example. As we observed, a coin can be a proxy for an on-line ad. Imagine the situation facing an agent who has a library of n ads that he/she can display on a particular web page. The agent is promised a payment of Rs. 10 every time

[1] See `https://en.wikipedia.org/wiki/Beta_distribution`.

a click is registered on the ad that is displayed. If the agent gets a total of T time slots for displaying an ad from among the n it has, which ad must it choose to display at each stage?

If, somehow, the agent already knew the click-through rate of each ad in its possession, its optimal strategy would be to always display an ad with the highest click-through rate. Such a strategy would maximise the expected number of clicks, and therefore the expected revenue. On the other hand, assume that our agent has no pre-existing knowledge. The only way the agent can gain knowledge is by displaying ads and observing whether a click occurs or not—a perfect setting for applying Bayesian inference! (More abstractly, our agent is faced with the problem of sampling the arms of a *stochastic multi-armed bandit*, so called for the name once given to slot machines in Las Vegas that would dispense rewards probabilistically. In general, rewards need not be binary; they can be real-valued, too.)

Starting with a blank slate, our agent necessarily has to *explore*: that is, display each ad a few times to get a good estimate of its click-through rate. At the same time, the agent cannot afford to explore too much (lest it displays unrewarding ads too many times); it needs to allocate more traffic to ads that seem to have a high chance of being clicked. In other words, the agent must *exploit* promising ads. There are several algorithms to optimise this "explore-exploit" trade-off. It can be shown that in order to perform well on every possible problem instance (a problem instance here is an element of $[0,1]^n$), an algorithm must necessarily display each suboptimal ad at least $\Omega(\log(T))$ times. Optimal algorithms are ones that display each suboptimal ad $O(\log(T))$ times.

Thompson Sampling [1, 2] is one such optimal algorithm, which is based on the principle of Bayesian inference. The algorithm begins with a uniform prior for each ad, and as the outcomes of displays are observed, the algorithm updates its beliefs exactly as described in Section 3. Thus, if an ad has been shown $h+t$ times, yielding h clicks and t no-clicks, the belief distribution corresponding to the ad is $\text{Beta}(h+1,\, t+1)$. At each stage, given the beliefs over each ad, the decision to make is which ad to display next. Thompson Sampling makes this decision randomly, but in accordance with the current belief distributions. Specifically, for each ad $i \in \{1, 2, \dots, n\}$ a sample $x_i \in [0,1]$ is drawn from the corresponding belief distribution. The ad selected to be displayed next is simply one that yielded the highest sample: namely $\text{argmax}_{i \in \{1,2,\dots,n\}} x_i$ (with ties broken arbitrarily).

The philosophy behind Thompson Sampling is as follows. At any stage, the agent has a belief distribution over the set of possible worlds. In our particular example, this belief distribution can be factored into belief distributions over the click-through rate of each ad. To decide which ad to show next, a world $(x_1, x_2, \dots, x_n)$ is sampled from the belief distribution. Having drawn the sample, the agent takes an action *as if* this world is real (which would mean $\text{argmax}_{i \in \{1,2,\dots,n\}} x_i$ is indeed an ad with the highest click-through rate). Interestingly, it took several decades after the algorithm was first introduced [1] to show that it is indeed optimal [3].

Thompson Sampling is not limited to the scenario that we have described above; it can also be applied in problems in which each world is a more complex object (such as an MDP). The underlying idea remains the same: (1) to maintain a belief distribution over the set of possible worlds, (2) to sample a world at each stage, and (3) to act optimally for that world.

5 Updating Beliefs in a Changing World

So far, we have assumed that there is a fixed world (in our example, a coin) that is sequentially generating evidence (outcomes of tosses). However, in general, the world itself can change with time. As an example, consider the problem of tracking a tiger that is wearing a radio collar. The "world" in this case would be the position of the tiger, and the periodic evidence it would provide is the strength of the radio signal. Crucially, the world itself could change over time (that is, the tiger could move around). However, we expect to have some knowledge about the process according to which the tiger moves.

The scenario above is best modeled using a Dynamic Bayes Net, an example of which is provided in Figure 5.

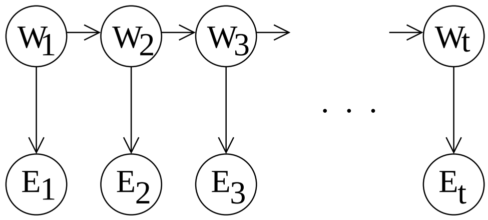

Figure 2: Dynamic Bayes Net with hidden variables $W_1, W_2, \ldots, W_t$ and observed variables $E_1, E_2, \ldots, E_t$.

As before, we maintain $Belief_t = P(W_t | E_1, E_2, \ldots, E_t)$ and seek to obtain $Belief_{t+1}$ based on $Belief_t$ and E_{t+1}. The only change from our previous working is that now, we also need to use $P(W_{t+1}|W_t)$—the transition model—which we assume is available based on our domain knowledge. For example, we may estimate that a tiger will be found uniformly at random within a 50km radius of its previous position. Our working yields:

$$
\begin{aligned}
&P(W_{t+1}|E_1, E_2, \ldots, E_{t+1}) \\
&= \frac{P(W_{t+1}, E_1, E_2, \ldots, E_{t+1})}{P(E_1, E_2, \ldots, E_{t+1})} \\
&= \frac{P(E_{t+1}|E_1, E_2, \ldots, E_t, W_{t+1}) P(E_1, E_2, \ldots, E_t, W_{t+1})}{P(E_1, E_2, \ldots, E_{t+1})} \\
&= \frac{P(E_{t+1}|W_{t+1}) P(E_1, E_2, \ldots, E_t, W_{t+1})}{P(E_1, E_2, \ldots, E_{t+1})} \\
&= \frac{P(E_{t+1}|W_{t+1})}{P(E_1, E_2, \ldots, E_{t+1})} \sum_{w \in W_t} P(E_1, E_2, \ldots, E_t, W_{t+1}, W_t = w) \\
&= \frac{P(E_{t+1}|W_{t+1})}{P(E_1, E_2, \ldots, E_{t+1})} \sum_{w \in W_t} P(W_{t+1}|E_1, E_2, \ldots, E_t, W_t = w) P(E_1, E_2, \ldots, E_t, W_t = w) \\
&= \frac{P(E_{t+1}|W_{t+1})}{P(E_1, E_2, \ldots, E_{t+1})} \sum_{w \in W_t} P(W_{t+1}|W_t = w) P(E_1, E_2, \ldots, E_t, W_t = w) \\
&= \frac{P(E_{t+1}|W_{t+1})}{P(E_1, E_2, \ldots, E_{t+1})} \sum_{w \in W_t} P(W_{t+1}|W_t = w) P(W_t = w|E_1, E_2, \ldots, E_t) P(E_1, E_2, \ldots, E_t) \\
&\propto P(E_{t+1}|W_{t+1}) \sum_{w \in W_t} P(W_{t+1}|W_t = w) P(W_t = w|E_1, E_2, \ldots, E_t).
\end{aligned}
$$

Thus, $Belief_{t+1}$ can be constructed from $Belief_t$, E_{t+1}, and $P(W_{t+1}|W_t)$.

References

[1] W.R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25:285–294, 1933.

[2] O. Chapelle and L. Li. An empirical evaluation of thompson sampling. In *NIPS*, pages 2249–2257. Curran Associates, 2011.

[3] Emilie Kaufmann, Nathaniel Korda, and Rémi Munos. Thompson sampling: An asymptotically optimal finite-time analysis. In *ALT 2012*, pages 199–213. Springer, 2012.